ARTICLES

Quantum conditional probability and hidden-variables models

James D. Malley*
Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892
(Received 22 December 1997; revised manuscript received 20 March 1998)

We discuss quantum conditional probability and its applications to deterministic hidden-variable models. We derive empirical tests corresponding to mathematical no-go proofs, providing rigorous statistical tests based on experimental outcomes. Evidently, it now possible to examine the statistical power of the empirical tests, and place confidence intervals on the parameters that precisely measure the departure of hidden-variable models from quantum experimental outcomes. Moreover, reinterpretation of well-known results in the light of quantum conditional probability provides other experimental demonstrations and no-go proofs: outcomes for the familiar Young double-slit experiment show that there are no deterministic hidden-variable models of the type considered by Kochen and Specker [J. Math. Mech. **17**, 59 (1967)] or Bell [Rev. Mod. Phys. **38**, 447 (1966)]. [S1050-2947(98)07908-6]

PACS number(s): 03.65.Bz, 03.65.Ca

I. INTRODUCTION

We introduce a class of theoretical proofs and empirical tests for quantum hidden-variables (HV) models. The family of models considered is the familiar one studied by Kochen and Specker [1], Bell [2], and many others. The main technical feature is the use of quantum conditional probability, defined for arbitrary pairs of noncommuting projectors. Another feature is that the empirical tests use as few as two projectors, and the proposed experiments are technically elementary. Also, single particles are all that is required, rather than measurements on correlated particle pairs as are used in connection with HV tests that depend on the Bell inequalities. The mathematical no-go proofs we obtain are exact, or ''inequality free,'' and hence may also be compared with the multiparticle, multioperator, exact no-go proofs introduced by Peres [3,4], Mermin [5,6], and Greenberger, Horne, Shimony, and Zeilinger [7].

In Sec. II, we give precise specifications for the class of HV models considered here, and provide a context by outlining the results of Fine and Teller [8] and Fine [9–11]. These results help display the connections between classes of HV models, and marginal and joint distributions for quantum outcomes. In Sec. III we define and discuss quantum conditional probability, a key fact being that its standard definition does not require the existence of a joint distribution for the observed and conditioning events; see Beltrametti and Cassinelli [12].

In Sec. IV we introduce a result from classical mathematical statistics linking conditional probability with the existence of (classical) joint distributions; see Arnold and Press [13]. These results are used to derive the rigorous statistical inference procedures that couple to the empirical tests we propose.

In Sec. V we pull these results together to derive theoretical and experimental methods to obtain no-go results for the HV models considered here. One interesting consequence is that the classic Young double-slit experiment, for example, provides an experimental test and a mathematical no-go proof for our class of HV models. Many other equally simple tests and proofs are also now possible.

Finally, in Sec. VI we discuss the proposed experimental tests of Peres [3,4].

II. THE CLASS OF HIDDEN-VARIABLES MODELS

We describe possible specifications for a hidden-variable HV model, such as are given by [1,2] and subsequently by many others. As a general reference for this topic one may use [12, Chap. 25], van Fraassen [14], or Bub [15].

Let $Q=Q(\mathcal{H},D,\mathcal{A})$ denote a quantum system with Hilbert space $\mathcal{H}$ (possibly of dim=2), a given arbitrary density operator D, and a family of observables $\mathcal{A}$.

Let $\Omega=\Omega(\Lambda,\sigma(\Lambda),\mu)$ denote a classical probability space, where Λ is a nonempty set, $\sigma(\Lambda)$ is a Boolean σ algebra of subsets of Λ, and μ is the probability measure on $\sigma(\Lambda)$.

As used in this paper, hidden-variable models for a quantum system in a given state D will make some or all of the following four assumptions.

HV (a): Given $\omega\in\Lambda$, $A\in\mathcal{A}$, there is a mapping f from the pair (ω,A) to real numbers; it is required that the value of $f(\omega,A)$ be an eigenvalue of A (the spectrum rule).

HV (b): For any two commuting operators A,B, the mapping f is such that $f(\omega,A+B)=f(\omega,A)+f(\omega,B)$ (the sum rule).

HV (c): The measure μ correctly returns the marginal probabilities for each observable A; for S a real Borel set, μ is such that

$$\Pr[A\varepsilon S]=\mathrm{tr}[DP_A(S)]=\int f(\omega,P_A(S))d\mu,$$

*Electronic address: jmalley@helix.nih.gov for $P_A(S)$ the projector associated with set S in the spectral resolution of A (the first-order margins rule).

HV (d): For any two commuting observables A and B, the measure μ correctly returns the joint probabilities; for S and T real Borel sets, μ is such that

$$\Pr[A\varepsilon S,B\varepsilon T]=\mathrm{tr}[DP_A(S)P_B(T)]$$
$$=\int f(\omega,P_A(S)P_B(T))d\mu$$

for $P_A(S),P_B(T)$ the projectors associated with sets S and T in the spectral resolutions of A and B, respectively (the second-order margins rule).

In HV (a), Λ is called the hidden *phase space* for the HV model, and *f* denotes the *valuation function* for the model. A complete specification for *f* must involve the density operator for the quantum system: $f(A,\omega)=f(\omega,A;D)\in\Re$, as we make no assumption about how *f* might transform under a change in density.

We write $A(\omega)$ to mean the classical random variable defined through the valuation assignment, where $A(\cdot)$ takes points in Λ to real numbers, so that $f(\omega,A;D)=A(\cdot)\in\Re$. Also, for projector X, and Borel set S and $\Re$, let

$$x=X^{-1}(1)=\{\omega\in\Lambda:X(\omega)=1\}. \tag{2.1}$$

Following [10], let us agree to call an HV model satisfying HV (a) and (c), a *weak hidden variables model*. A model satisfying conditions HV (a), (c), and (d) will be termed a *deterministic hidden variables model*. In words, for a weak HV model the valuation function provides a map from all operators to a space of classical random variables. It need not be pointwise linear on each operator: the valuation function assignment only needs to be linear on average for single observables, and it must return the correct marginal probabilities for each observable. Also, given HV (c), the valuation assignment is such that its integration over the whole space is linear, but in general this is strictly weaker than requiring HV (b), since the integral equality in HV (c) is not equivalent to assuming

$$\int_\Delta f(\omega,X+Y)d\mu=\int_\Delta f(\omega,X)d\mu+\int_\Delta f(\omega,Y)d\mu \tag{2.2}$$

for Δ an arbitrary Borel set in $\sigma(\Lambda)$ and projectors X,Y. Given HV (a) and HV (c), it follows that HV (b) is obtained if and only if the integration above holds when taken over all possible Δ.

All our results below will be obtained under the set of conditions HV (a), (c), and (d): we fix our class of interest as that of the deterministic HV models. Fine [10] has demonstrated that this is exactly the model considered, for example, by Kochen and Specker [1]. Moreover, we show how to replace conditions (b) and (c) with a single equivalent premise on conditional probability and upon this our no-go proofs and tests will be based.

We recall that useful equivalent model conditions are available. Fine [9,10] proves that in the presence of HV (a), condition HV (b) is equivalent to either of the following.

HV (b_1): For any two commuting operators A,B, the valuation is such that $(AB)(\omega)=A(\omega)\cdot B(\omega)$ (the product rule).

HV (b_2): For any Borel measurable function g, and any operator A, the valuation is such that $[g(A)](\omega)=g[A(\omega)]$ (the Borel function rule).

It is important to note that condition HV (a) is an assignment of eigenvalues for each operator that is made once for all operators (commuting or not), but the sum rule HV (b) [equivalently (b_1) or (b_2)] applies to commuting operators only.

Some additional context for the family of HV models discussed here is now provided. Consider first the small but important models of Bell [2] and Kochen and Specker [1]. These both assume dim $\mathcal{H}=2$, and are examples for which linearity or the factoring rules [HV (b), (b_1), or (b_2)] are not assumed to hold. See, for example, the discussion in [12], pp. 268–271, where it is shown that the sum rule, condition HV (b), is violated for these two models. Otherwise expressed, these authors provided two distinct, probability measures on the lattice of projectors (in the spin-$\frac{1}{2}$ space of dim $\mathcal{H}=2$) that do not extend linearly from the space of projectors to space of all bounded operators. On the other hand, as discussed by Bell [1], an application of Gleason's theorem mathematically rules out the family of HV models of the type we consider here, in the case dim $\mathcal{H}\geqslant 3$.

Next, given that HV (a) and HV (c) hold, Fine [9] shows that HV (b_2) is equivalent to HV (d).

Theorem (Fine [9]). Given the spectrum rule and the first-order margins rule, the second-order margins rule is equivalent to the function rule.

One of the excellent features of Fine's result is that it focuses attention on experiment and observation, rather than on apparently more esoteric features of a valuation function. That is, experimentally determinable first- and second-order margins alone are sufficient to fix important pointwise properties of the valuation. Moreover, from Fine [10], p. 292: "the idea of deterministic hidden variables is just the idea of a suitable joint probability function," *on the phase space*, and that correctly returns the observed marginal and joint probabilities.

Regarding other possible demands one may place on an HV model, we note that a deterministic HV model is not contextual, as it makes its valuation assignment once for all observables (at each point in the phase space). It is also not contextual in the sense considered by Gudder [16], where the phase space probability measure is permitted to vary, depending on each set of commuting operators under consideration. As a deterministic HV model makes an assignment that factors for any commuting pair (the product rule), it is also a *local* model; see [12], p. 274.

We turn now to the definition and properties of quantum conditional probability.

III. CLASSICAL AND QUANTUM CONDITIONAL PROBABILITY

Following [12], Chap. 26, for two projectors A and B, not necessarily commuting, on a quantum system in state D, *quantum conditional probability* is defined to be

$$\Pr_D[A|B]=\mathrm{tr}[DBAB]/\mathrm{tr}[DB] \quad \text{when } \mathrm{tr}[DB]\neq 0. \tag{3.1}$$

Note that

$$\Pr_D[A|B]=\Pr_{D(B)}[A], \quad \text{where } D(B)=BDB/\mathrm{tr}[DB]. \tag{3.2}$$

Some remarks are in order. First, note especially that while in classical probability the definition of conditional probability for classical outcomes is made by means of a joint probability (or is itself made a basic postulate, and a joint probability is defined from it), it is the case that quantum conditional probability does not require a joint distribution, and is defined for every projector pair A,B, commuting or not.

Alternative approaches to evaluating conditional or unconditional probabilities for quantum events might choose to change the range space for the probability measure. Here, one change could be from a real valued, positive measure to a signed or possibly complex measure.

A recent example of this approach is from Scully, Walther, and Schleich [17], using a probability primitive defined on projectors that can take on negative values. In doing this they invoke an idea due to Feynman, and report interpretive simplifications of the EPR problem. Their construction is also used to reinterpret a Young double-slit experiment, in the form of a micromaster, *Welcher-Weg* detector. This may be compared with the reevaluation described below at the end of Sec. V.

Scully *et al.* [17] do not use their revised notion of probability to derive hidden-variable model results. The systematic use of conditional quantum probability, on the other hand, will lead to new conclusions about hidden-variables models, as well as to fruitful reinterpretations of familiar results. Just how their results could be derived from ours is briefly considered at the end of Sec. V.

Next, note that the density operator $D(B)$ is the quantum state of the system given that the event associated with the projector B has occurred; it is the state of the system after the projector B has been applied to the system. The connection between conditional probability and sequential measurements is discussed in many places, including Bohm [18], pp. 67–74 and Helstrom [19], pp. 65–69.

The connections between sequential measurements and HV models are, however, still controversial. Some details of this discussion appear in Bub ([20,21,15], Chaps. 2 and 3), Freedman and Wigner [22], Clauser [23,24], and Wigner [25]. To illustrate the problem, and to resolve some of the discussion, we introduce the following notation: for S, any set in $\sigma(\Lambda)$, let $\mu[S]=\int_S d\mu$, and for any projector X, let $x=X^{-1}(1)=\{\omega\in\Lambda:X(\omega)=1\}$ [as in Eq. (2.1)].

Now, for any two projectors A,B (not necessarily commuting), and given a deterministic HV model we can always define

$$\mu[a|b]=\mu[a\cap b]/\mu[b] \quad \text{when } \mu[b]\neq 0. \tag{3.3}$$

One of the central questions for conditional probability and sequential measurements is: when does $\mu[a|b]=\Pr_D[A|B]$? To study this, we introduce a new HV model assumption:

HV (e): For any two projectors A,B (not necessarily commuting),

$$\mu[a|b]=\mu[a\cap b]/\mu[b]=\Pr_D[A|B]=\mathrm{tr}[DBAB]/\mathrm{tr}[DB]$$

$$(\text{the conditional probability rule}).$$

Also, for any two projectors A,B, write $A\leqslant B$ to mean $AB=BA=A$.

Lemma. Given the spectrum and the product rules, if A,B are two projectors such that $A\leqslant B$, then $a\cap b=a$.

Proof. By the product rule, $(AB)(\omega)=A(\omega)B(\omega)=A(\omega)$. Using the spectrum rule, if it is given that $A(\omega)=1$, then necessarily $B(\omega)=1$, so that $\{\omega|A(\omega)=1, \text{ and } B(\omega)=1\}\supseteq\{\omega|A(\omega)=1\}$. Since it is always the case that $\{\omega|A(\omega)=1, \text{ and } B(\omega)=1\}\subseteq\{\omega|A(\omega)=1\}$, the result follows.

We now prove our main result on the relationship between HV models and the conditional probability rule.

Theorem 1. Assume dim $\mathcal{H}\geqslant 3$. The spectrum rule and the first- and second-order margin rules together imply that the conditional probability rule holds. Conversely, given that the spectrum rule is valid, if the conditional probability rule holds then the first- and second-order margin rules hold when these are restricted to the lattice of projectors.

Proof. Assume that there is a deterministic HV model, so that the first- and second-order margin rules are valid. For two events, a and b in $\sigma(\Lambda)$, the classical definition for conditional probability applied to sets in phase space is $\mu[a|b]=\mu[a\cap b]/\mu[b]$, and one checks that this serves to formally define a consistent probability measure on the lattice of all projectors (where A and B need not commute). In particular, for projectors A_1 and A_2 such that $A_1A_2=A_2A_1=0$, and $A=A_1+A_2$, by using the spectrum and margin rules, one can show that $\mu[a|b]=\mu[a_1|b]+\mu[a_2|b]$.

Next note that $A\leqslant B$ implies $\Pr_D[A|B]=\Pr_D[A]/\Pr_D[B]$, and by the Lemma above we also have $\mu[a|b]=\mu[a]/\mu[b]$. Hence using the first-order margins rule we find $\mu[a|b]=\Pr_D[A|B]$, whenever $A\leqslant B$. Using Gleason's theorem this implies that $\mu[a|b]=\Pr_D[A|B]$ for any two projectors, as required. A detailed proof of this important fact appears in [12], p. 288.

The proof of the converse, that the conditional probability rule implies the first- and second-order margin rules, when restricted to projectors, is straightforward and is omitted.

A proof of *Theorem 1* in the special case that the conditioning projector is one-dimensional is possible using a considerably weaker form of Gleason's theorem. Thus, Gudder ([26], p. 129, corollary 5.17) obtains the following by elementary vector space methods.

MicroGleason. For dim $\mathcal{H}\geqslant 3$, if a probability measure m on the lattice of projectors assigns probability 1 to any one-dimensional projector B, then it must be such that $m(A)=\mathrm{tr}[BA]$, for all projectors A.

When no-go proofs or experimental tests, obtained below, involve only pairs of one-dimensional projectors, only microGleason is needed. The other proofs and experimental tests for HV models proposed below evidently do require the uniqueness feature of the original Gleason theorem. While this compromises the simplicity of the mathematical form of these no-go proofs, compared with other results not using Gleason (Kochen and Specker, or Bell, for example), it does not effect the merit of the empirical tests we propose for deterministic HV models.

In order to arrive at some of these experimental tests and proofs of HV models, we need another fact from the calculus for quantum conditional probability. That is, [12], Chap. 26 shows how quantum conditioning on an orthogonal sum of commuting projectors yields interference terms, sharply distinguishing it from classical conditioning. For simplicity, consider a sum of two commuting, orthogonal projectors:

$$Q=Q_1+Q_2 \quad \text{for projectors} \quad Q_1,Q_2$$

$$\text{with} \quad Q_1Q_2=Q_1Q_2=0.$$

Assume the quantum system is the pure state $D=|\psi\rangle\langle\psi|$, with $\psi\in\mathcal{H}$. Then

$$\Pr_D[P|Q]=\left\langle \frac{Q\psi}{\|Q\psi\|}\right| P \left. \frac{Q\psi}{\|Q\psi\|}\right\rangle. \tag{3.4}$$

Letting $\varphi_i=Q_i\psi/\|Q_i\psi\|\neq 0$, for $i=1,2$,

$$\Pr_D[P|Q]=\left(\frac{\|Q_1\psi\|}{\|Q\psi\|}\right)^2 \Pr_D[P|Q_1]+\left(\frac{\|Q_2\psi\|}{\|Q\psi\|}\right)^2 \Pr_D[P|Q_2] + \frac{\|Q_1\psi\|\cdot\|Q_2\psi\|}{\|Q\psi\|^2}\,\mathrm{Re}\langle\varphi_1|P\varphi_2\rangle. \tag{3.5}$$

The last term in the conditional probability equation above represents quantum interference. This is a nonclassical feature of quantum conditional probability, and if the interference term is nonzero it ought to be experimentally observed. In other words, quantum conditioning, when conditioning is taken over orthogonal sums, does not in general return a classical convex mixture over the components in the sum.

We more fully exhibit the connection of this result with properties of a deterministic HV model. Given the existence of such an HV model, we can use a classical probability calculus result on conditioning over sums. For arbitrary projectors A, B, U, and V, let the sets a, b, u, and v be defined as in Eq. (2.1), and assume that $B=U+V$, with $UV=VU=0$. One can check that $b=u\cup v$. Then, using standard classical probability rules (apart from any special assumptions for an HV model) we have

$$\begin{aligned}\mu[a|b]&=\mu[a\cap b]/\mu[b]=\mu[a\cap(u\cup v)]/\mu[u\cup v]\\&=\mu[(a\cap u)\cup(a\cap v)]/\mu[u\cup v]\\&=\{\mu[a\cap u]+\mu[a\cap v]\}/\mu[u\cup v]\\&=\{\mu[u]/\mu[u\cup v]\}\mu[a|u]\\&\quad+\{\mu[v]/\mu[u\cup v]\}\mu[a|v].\end{aligned}$$

Hence

$$\mu[a|b]=p\cdot\mu[a|u]+q\cdot\mu[a|v], \tag{3.6}$$

where $p=\mu[u]/\mu[b]$, $q=\mu[v]/\mu[b]$, and $p+q=1$.

In light of *Theorem 1*, assumption of a deterministic HV model must imply that conditioning over a disjoint sum (equivalently, over an orthogonal sum of projectors) results in a classical convex mixture over separate conditionals. That is, the quantum interference term must vanish if the deterministic HV model is valid. Note also that all terms in the last equation correspond to observable conditional or marginal events, and may be evaluated from experiment.

As a useful physical example of this conditioning process, one that we can use later in our no-go proofs and empirical tests, consider a spin-1 particle and two Stern-Gerlach devices that sequentially separate the possible values $-1,0,+1$ of the spin component along the x and the y axis, respectively. Let Q be the event where ''the x component of the spin is -1 or $+1$,'' and let P be the event where ''the y component of the spin is $+1$.''

In order to make the second measurement conditional on Q, rather than on just Q_1 or just Q_2, we must assume that the output of the first device is coherently recombined before being sent to the second device. In particular, no determination is being made by the first device of either ''the x component of the spin is -1,'' or ''the x component of the spin is $+1$.''

On the other hand, we can obtain a conditional probability that conforms more closely to a classical mixture by modifying the experimental arrangement: use the first device to measure ''x component of the spin is $+1$,'' separately from ''x component of the spin is -1.'' This is done by directing the two spin outcomes separately to the second device, that is, by not coherently recombining them before presentation to the second device. In this second setup we observe the spin x components, and the quantum conditional probability reduces to a more classical, convex probability mixture represented by the first two terms in the equation above.

The next section deals with statistical inference issues that support the empirical tests we propose.

IV. STATISTICAL INFERENCE ISSUES

A. A result from classical statistics

Suppose we are given two discrete, classical random variables X and Y, and we observe sets of conditional outcomes:

$$\Pr[X=x|Y=y] \quad \text{and} \quad \Pr[Y=y|X=x].$$

We ask if these two conditional distribution are *compatible*, in the sense that they derive from a single, unobserved joint distribution $\Pr[XY]$, such that

$$\begin{aligned}\Pr[X=x,Y=y]&=\Pr[X=x|Y=y]\cdot\Pr[Y=y]\\&=\Pr[Y=y|X=x]\cdot\Pr[X=x].\end{aligned} \tag{4.1}$$

Arnold and Press [13] describe several classical (nonquantum) data analysis problems for which stated conditional distributions are, or are not, compatible. They present results that apply to both discrete or continuous measurements, as well as for higher-dimensional cases. And while they do not provide a statistical test for compatibility, they do give simple necessary and sufficient formal conditions from which such tests can be derived. This we do below, after describing the original results.

A conceptual obstacle may be present here: we are accustomed to beginning with a pair of random variables X,Y and then making the standard existence statement about a joint distribution, which classically always exists. The situation discussed here and in [13] is different: start with conditional distributions of a *given, explicit form* and then find, if possible, a joint distribution consistent with these conditionals.

To motivate their result, [13] also gives several illustrations from the classical statistical literature showing when no joint distribution can possibly exist if it is required to correctly return the stated conditional distributions. We describe one of these. Thus, suppose that the conditional distribution of X, given $Y=y$, is a Gaussian with expectation$=ay+by^2$ and variance $f(y)$, while the conditional distribution of Y, given $X=x$, is Gaussian with expectation$=cx+dx^3$ and variance $g(x)$, for constants a, b, c, and d. Then one can show that X and Y have a joint distribution if and only if $d=0$. The critical requirement is that any proposed joint distribution must return the correct, stated conditional distributions. Variables X and Y can have no joint distribution if the conditionals are as given, unless $d=0$.

We next take up the solution given in [13] to this problem of connecting given conditional distributions to a single joint distribution.

B. A formal existence condition

For simplicity, consider only the discrete case, with $1\leqslant i\leqslant I$, $1\leqslant j\leqslant J$, and let

$$a_{ij}=\Pr[X=x_i|Y=y_j] \quad \text{and} \quad b_{ij}=\Pr[Y=y_j|X=x_i]. \tag{4.2}$$

Assume that $a_{ij}>0$, $b_{ij}>0$, for all i and j.

[13] proves that the two conditional distributions are compatible if and only if there exist two vectors of constants

$$\mathbf{r}=(r_1,r_2\ldots,r_I) \quad \text{and} \quad \mathbf{s}=(s_1,s_2,\ldots,s_J),$$

with all entries positive, such that

$$a_{ij}=r_is_jb_{ij} \quad \text{for all} \quad i,j. \tag{4.3}$$

Equivalently we may use the condition

$$a_{ij}=dr_is_jb_{ij} \quad \text{for all} \quad i,j, \quad \text{and some real constant} \quad d>0, \tag{4.4}$$

and let us agree to call this the AP condition.

Given the positivity conditions above, [13] also shows that the joint distribution is unique, if it exists, and given $\mathbf{s}$, the joint distribution can be formally found as follows.

First verify that defining

$$\mathbf{t}=(t_1,t_2\ldots,t_I), \quad \text{where} \quad t_1=r_1/\Sigma r_i \quad \text{all} \quad i, \tag{4.5}$$

yields a consistent marginal for X, such that

$$\mathbf{t}=(\Pr[X=x_1],\Pr[X=x_2],\ldots,\Pr[X=x_I]). \tag{4.6}$$

Then $\Pr[XY]$ is found from the IJ equations:

$$\Pr[X=x_i, \quad Y=y_j]=\Pr[Y=y_j|X=x_i]\cdot\Pr[X=x_i]=a_{ij}t_i. \tag{4.7}$$

As an illustration, consider the following special case.

Lemma. If $\Pr[X=x_i|Y=y_j]=\Pr[Y=y_j|X=x_i]$, for all i,j, then the AP condition holds, with $\Pr[Y=y_j]=\Pr[X=x_i]$ for all i and j, and

$$\mathbf{t}=(1/I)(1,\ldots,1).$$

Proof. If the conditional probabilities are all equal, then the AP condition certainly holds with $r_i=s_j$ for all i and j, and a joint distribution exists for X and Y. Hence

$$\Pr[X=x_i,Y=y_j]=\Pr[X=x_i|Y=y_j]\cdot\Pr(Y=y_j)$$
$$=\Pr[Y=y_j|X=x_i]\cdot\Pr[X=x_i].$$

Equality of the conditionals implies that $\Pr[Y=y_j]=\Pr[X=x_i]$ for all i and j. The result for $\mathbf{t}$ is immediate.

More generally, if the AP condition holds, some formal solution to the compatibility problem must exist: there must be a unique joint distribution function that returns the given conditional distributions. We observe, though, that the AP condition by itself may not validate the existence of a joint distribution that returns a specific set of hypothesized first-order marginal distributions. More precisely, we argue that a logically complete procedure for testing the existence of a joint distribution must ensure that marginals derived from the vector $\mathbf{t}$ defined above match a given set of marginals. This is an additional question not considered in [13], but one that we take up in the next subsection.

We have one further issue to address, also not considered in [13]. It is that the AP condition is not itself a statistical test, applicable to data and from which an inference can be drawn. It is straightforward to provide a statistical procedure to check for matching marginals, as well as for the AP condition itself, and this is discussed in the next subsection.

Finally, note that a formal calculation or an empirical test, for a pair of random variables, may reject the AP condition, in which case a test for the marginals is not needed.

C. A statistical test for the AP condition

We use standard results for so-called log-linear models, which apply to the statistical analysis of multiway tables of discrete, counted data; see, for example, Agresti [27] or Bishop, Fienberg, and Holland [28]. Data collection is first described, using two noncommuting projectors, and then the statistical procedure is given.

Consider a spin-1 system, obtaining observations on a single particle, given that the system is in state D. Suppose a measurement on the same particle is then made using, in turn, one and then the other of any two noncommuting one-dimensional projectors X and Y.

Outcomes of X will occur according to the marginal probability $\Pr_D[X]$. A subsequent Y measurement made on the same particle will occur according to the conditional probability $\Pr_D[Y|X]$. Similarly, measurements made first on Y, followed by one on X, will have corresponding probabilities $\Pr_D[Y]$ and $\Pr_D[X|Y]$, respectively.

If we first randomly select X (with probability $\frac{1}{2}$) for the first measurement, and allow (filter) only outcomes $X=1$ to be sent to the second spin measuring device, we find that $Y=1$ outcomes from the second device will be registered with probability $(\frac{1}{2})\Pr[Y=1|X=1]$.

More generally, randomly selecting the first detector filter, filtering the first outcomes, and then observing spin at the second detector will produce counts n_{ijk} in a three-way contingency table, with each cell probability, p_{ijk}, as just described.

One standard assumption made about such data collection (in any multiway contingency table) is that the individual cell counts have a Poisson distribution, which is also the usual one for photon counts. Other measurement and sampling schemes may be used instead for which the statistical methods proposed here are still valid asymptotically. It is an important feature of these Poisson cell counts that if we make a total of N observations, then the cell probabilities conditional on N will have a multinomial distribution, and then $Np_{ijk}=m_{ijk}$, the expected value for cell (i,j,k). In what follows, we assume a multinomial distribution for the cell counts.

We can write a full or saturated model for all the cell counts in the three-way table of counts in terms of expected values, m_{ijk}, for the cells, where a saturated model contains as many independent terms as there are data cells. Let $Z=k$ $(k=1,2)$ denote the outcome of randomly choosing to measure X (or Y) first, to be followed by a conditional measurement of Y (or X). Then the model has the form

$$\ln(m_{ijk})=u+u_i^X+u_j^Y+u_k^Z+u_{ij}^{XY}+u_{ik}^{XZ}+u_{jk}^{YZ}+u_{ijk}^{XYZ} \tag{4.8}$$

for parameters in the variables u, which all become identifiable given the constraints

$$\Sigma u_i=\Sigma u_j=\Sigma u_k=\Sigma u_{ij}=\Sigma u_{ik}=\Sigma u_{jk}=\Sigma u_{ijk}=0. \tag{4.9}$$

Using the sampling scheme described above, with $Np_{ijk}=m_{ijk}$, it follows that

$$\ln(a_{ij})=\ln(m_{ij1})-\ln(1/2)-\ln N, \tag{4.10}$$

$$\ln(b_{ij})=\ln(m_{ij2})-\ln(1/2)-\ln N, \tag{4.11}$$

and

$$\ln(a_{ij}/b_{ij})=\ln(m_{ij1}/m_{ij2}). \tag{4.12}$$

It is straightforward to verify the following.

Theorem 2. The AP condition is obtained if and only if the three-way interaction term u_{ijk}^{XYZ} is identically zero, for all i, j, and k.

In order to experimentally test that the three-way interaction is zero, we can use a standard, large sample χ^2 analysis, or essentially equivalently, a maximum likelihood estimation of the expected cell values, followed by a likelihood ratio test. The test we propose here is also known as the goodness of fit for a *homogeneous association model*; see [27], pp. 151 and 152.

As a practical matter, the maximum likelihood estimate does not have a direct or closed-form solution for a model with just the three-way interaction equal to zero. However, a simple iterative procedure exists, or one could use any of a number of commercial software systems, such as S-PLUS, SAS or STATXACT.

In the case of two projectors (not necessarily one-dimensional) we have $1\leq i,j\leq 2$, and the test for no three-way interaction reduces to the Breslow-Day statistic, which is widely used in clinical trials and epidemiology. Using STATXACT the test can be run using a nonparametric version, called Zelen's exact test; see [27], p. 66. This procedure makes no assumptions regarding the asymptotic distribution of the test statistic.

To test for correct marginal distributions for X (or Y), we use knowledge of the quantum system to formally calculate the conditional probabilities, which form the values for a_{ij} and b_{ij}. Under the null hypothesis that a suitable joint distribution holds, the AP condition must hold as well, and we can formally solve for the unique vector $\mathbf{t}$. This formal solution is in turn compared to the vector of observed marginals for X, by using a χ^2 multinomial test (for example) that compares observed counts against expected counts.

This completes the description of our two-step, statistical procedure for testing for the existence of a joint distribution, compatible with the given, observed conditional and marginal probabilities.

Other classical statistical procedures could be used in place of either the first or second step of the two step procedure we have just outlined. In particular, maximum entropy estimation, classical Bayes, empirical Bayes, and others; see [28], Chap. 10.

V. MATHEMATICAL NO-GO PROOFS AND EMPIRICAL TESTS

We are now finally in a position to properly interpret and apply our results to the question of the existence of deterministic HV models. We do this by joining *Theorem 1* with *Theorem 2*.

Theorem 3. Suppose X and Y are two observables for a quantum system, with associated projectors $\{X_i\}$ and $\{Y_j\}$ derived from their spectral decompositions. If there is a deterministic HV model for the quantum system, then the AP condition applied to the projectors is valid.

Proof. Assume that a deterministic HV model for the system exists. Then there must exist a phase space with a classical probability measure μ such that the conditional probabilities match: $\Pr_D[X_i|Y_j]=\mu[x_i|y_j]$, and such that the first-order margins also match: $\Pr_D[X_i]=\mu[x_i]$, and $\Pr_D[Y_j]=\mu[y_j]$.

Since $\mu[x_i|y_j]\cdot\mu[y_j]=\mu[y_j|x_i]\cdot\mu[x_i]$, it follows that

$$\Pr_D[X_i|Y_j]\cdot\Pr_D[Y_j]=\Pr_D[Y_j|X_i]\cdot\Pr_D[X_i]. \tag{5.1}$$

But this means that the AP condition must hold, with $r_i=\Pr_D[X_i]$, and $s_j=\Pr_D[Y_j]^{-1}$ as required.

Note that X and Y need not be commuting, and that the AP condition is applied to the *family* of conditional probabilities $\Pr_D[X_i|Y_j],\Pr_D[Y_j|X_i]$. The validity of the conditional probability rule ratifies the connection between observables and phase space, so for deterministic HV models it is the case that $\mu[a|b]=\Pr_D[A|B]$.

Though we do not need it here, it is possible to prove a kind of converse to *Theorem 3*. This is accomplished by showing that the AP condition for a pair of observables generates a classical measure space and a corresponding reduced deterministic HV model, one that is restricted to that pair of observables.

Most importantly, we note that the statistical procedures just described allow for exact and empirical evaluations of the power of the statistical tests. This requires evaluation of the probability that we reject the null hypothesis of a deterministic HV model, when in fact it is correct to do so. In the context of the AP condition, by specifying a set of nonzero interaction terms, we can, for a given sample size, calculate (or estimate) the probability that we will correctly reject the null model. For this same range of alternative models, and for a stated statistical power, we can also evaluate the sample size needed to correctly reject the null model.

On the other hand, such an analysis of the statistical properties of an empirical test does not appear to have been discussed for no-go tests of the form considered by Aspect *et al.* [29] and many others: if we suppose that quantum mechanics (the alternative model) holds rather than a deterministic one (the null model), then those procedures evidently do not clearly provide a conventional sample space and probability measure by which probabilities for *all* outcomes (those under the null or under the alternative) can be calculated. In particular, the acceptance probability under the alternative model (quantum theory) could be rather small, thereby diminishing the value of this procedure that draws inferences from the experimental outcomes. A closely related problem is that a given empirical test can be too frequently rejecting a deterministic HV model in favor of quantum theory. A discussion of these issues and the so-called type I and type II errors in statistical inference appears, for example, in Devore ([30], pp. 100–106).

We illustrate our results for a spin-1 system. Let two one-dimensional projectors be defined by X and Y, for which $\mathrm{tr}[D(XYX)]\neq \mathrm{tr}[D(YXY)]$.

If a deterministic HV model is valid, then using just microGleason [see the discussion following *Theorem 1*] the conditional probability rule must hold. Therefore $\mu[x|y]\cdot\mu[y]=\mu[x\cap y]=\mu[y|x]\cdot\mu[x]$, where $\mu[x|y]=\mathrm{Pr}_D[X|Y]$, and $\mu[y|x]=\mathrm{Pr}_D[Y|X]$. Simplifying, it must be the case that $\mathrm{tr}[DYXY]=\mathrm{tr}[DXYX]$, which is a contradiction.

We have, therefore, mathematically ruled out a deterministic HV model for the spin-1 system, using microGleason and just two projectors. Also, the proof uses only single particles and is inequality-free. Acardi [31] gives a proof very similar to this, but effectively *assumes* that the conditional probability rule holds. In Sec. IV C we described the data collection process for an empirical test corresponding to this mathematical result.

Consider next a structural feature of this test. Define

$$\lambda(X;Y)=\frac{\mathrm{Pr}_D[X|Y]\mathrm{Pr}_D[(1-X)|(1-Y)]}{\mathrm{Pr}_D[X|(1-Y)]\mathrm{Pr}_D[(1-X)|Y]},$$

$$\lambda(Y;X)=\frac{\mathrm{Pr}_D[Y|X]\mathrm{Pr}_D[(1-Y)|(1-X)]}{\mathrm{Pr}_D[(1-Y)|X]\mathrm{Pr}_D[Y|(1-X)]}. \qquad (5.2)$$

Now, an exact measure of the divergence of a deterministic HV model from quantum theory are the sizes of the u terms in the log-linear model that correspond to the three-way interaction. Using the natural constraints on the log-linear model, one shows there is only one absolute value, $=d$ say, for these u terms (see [28], p. 34), and that

$$d=(1/8)\ln[\lambda(X;Y)/\lambda(Y;X)]. \qquad (5.3)$$

This single expression contains all information about the divergence between the two models. After simplifying we get

$$d=\tfrac{1}{8}\ \ln\left[\frac{1-\mathrm{Pr}_D[X|(1-Y)]^{-1}}{1-\mathrm{Pr}_D[Y|(1-X)]^{-1}}\right]. \qquad (5.4)$$

Using the fact that maximum likelihood estimates of the u terms have large-sample Gaussian distributions, we can generate a confidence interval for d, therefore finding real constants ℓ and u such that the interval $[\ell,u]$ contains the true value of d with, say, 95% probability. Making this straightforward is the fact that most statistical software systems automatically report the large-sample standard deviation for the estimates of the u terms. On the other hand, for any given quantum system (one for which d is presumably not zero), we can explicitly calculate d, and thus evaluate the power of the test to distinguish between the two models; see [28], pp. 494–500 for complete details. This power could in principle be maximized by suitable selection of the projectors, given that the null model is rejected with a stated type I error.

We turn next to alternative proofs and empirical tests for deterministic HV models, where we use a different feature of quantum conditional probability.

If a joint phase space distribution exists for P,Q and the two components of Q, then the conditional probability over the sum must be, as shown earlier, a classical convex mixture of conditional probabilities. As discussed above, one consequence of this is that the quantum interference term must vanish. However, the existence of the interference term in the quantum conditional probability for the projectors P and Q given earlier is a strong suggestion that no deterministic HV model can apply to these operators.

All that is required to obtain a rigorous mathematical no-go proof is a set of projectors P and Q, with $Q=Q_1+Q_2$ and $Q_1Q_2=Q_2Q_1=0$, such that none of the factors in the interference term is exactly zero. The spin-1 case outlined above at the end of Sec. III is one such example, so it furnishes an alternative no-go proof for a deterministic HV model in the case dim $\mathcal{H}=3$. It uses only three projectors.

Parallel to these mathematical proofs, experimental tests can be arranged. We evaluate the interference term by performing the two experiments described above, the first where the output beams from the first device are coherently recombined, and the second experiment where they are not recombined. The numerical difference in the outputs from these two setups is precisely the quantum interference term. The empirical test is, therefore, the determination that this observed difference effectively vanishes, that is, is below estimated noise levels.

Beltrametti and Cassinelli ([12], Chap. 26) also describe the consequences for quantum conditioning in the case of the familiar Young double-slit experiment, though they do not do so in the context of hidden-variables models. They correctly include the free-evolution operator for the particle in transit and derive the density for particle detection at the collection screen in the separate cases of one or the other slit closed, and the case of both slits open. The open-closed status of the slits acts as the probability conditioning events.

The occurrence of the quantum superposition term, in the case of both slits open, exactly corresponds to the nonclassical probability interference term we considered earlier in the spin-1 system. Once again, this quantum conditioning effect generates a mathematical, no-go proof as well as an empirical test.

Because of quantum conditional probability, and *Theorem 3*, the double-slit experiment is therefore a classroom-level demonstration of no hidden-variables models. And the numerical difference between the wave and particle models for the particle is therefore reexpressible as the difference between classical and quantum conditional probability.

Finally, we consider how a suggestion of Feynman on negative ''probability'' may be derivable from quantum conditional probability. First recall that Scully *et al.* [17] review and advance the notion of probability for quantum events that can assume negative values.

We speculate that this approach may be derivable from our own. That is, for any two events *a* and *b*, in a classical probability space, it is always the case that

$$\mu[a]=\mu[a|b]\cdot\mu[b]+\mu[a|\text{not } b]\cdot\mu[\text{not } b]. \quad (5.5)$$

Consider replacing the classical conditionals above by the correct quantum conditionals, those of the form $\mathrm{Pr}_D[A|B]$, and assume the rule applies to quantum events. If the quantum conditionals now have their quantum interference terms deleted, then $\mu[a]$ may assume negative values. In this way it might also be possible to embed Feynman's original idea in a context that is rigorous and consistent. We pursue this approach elsewhere.

VI. DISCUSSION

Alternative empirical tests for no-hidden-variables models have been suggested by Peres [3,4]. While not being stated in terms that directly relate to the definitions given above, we will assume that the models considered by Peres invoke at least some of the same assumptions considered here. In particular, that there is a valuation function assignment for all projectors satisfying the rules for a deterministic HV model: the spectrum rule and the Borel function rule hold.

First, in Peres [3], an inequality-free no-go proof is obtained using a square array of nine operators, in the presence of condition HV (b_1), the product rule above. This leads to a valid mathematical no-go proof for a deterministic HV model: see also the review and discussion paper Mermin [6]. Peres also points out that each set of commuting projectors in the array could in principle be separately measured using generalized Stern-Gerlach devices. This author does not know if such an experiment has been carried out.

There seems to be, however, a deeper issue to resolve about the interpretation of such an experiment. That is, for an empirical test to directly correspond to the theoretical situation of the mathematical proof, each distinct measurement on the subsets of commuting projectors would have to be done with the phase space variable ω in the same, fixed but unknown setting. Allowing different values for ω across the incompatible experiments nullifies the conditions of the mathematical proof, while in an experimental setting it is not clear if activity in the hidden phase space can be controlled as required.

In particular, replication of any single experiment does not require fixing ω, under any HV model assumption so far made or considered. Some additional connection between phase space, valuation function, observation, and prediction would seem to be required.

Second, in Peres [4], a different approach is taken. Thus, let u,v be a pair of orthonormal state vectors, and let x,y be another orthonormal pair, with u,x and v,y noncommuting. Assume that these state vectors are arranged such that the corresponding projectors satisfy the equation

$$P_u+P_v=P_x+P_y. \quad (6.1)$$

Taking expectations for the quantum system prepared in a given state, standard quantum theory predicts that the same real number should result for both sides of this equation. Peres argues that this equality is in fact an assumption, and in principle a testable hypothesis. Using spin-1 systems and generalized Stern-Gerlach devices, an experimental scheme is displayed to evaluate this presumed equality. We are not aware if the scheme has been undertaken.

At this point, it is not clear to this reader if the deterministic HV model assumptions we have used are in fact those being tested by the experiment suggested by Peres [4]. If we assume that these assumptions are exactly those for a deterministic HV model, then under the sum rule [condition HV (b)], a valuation function would assign the same number to these operator sums, so that expectations taken *over the phase space* should return the same values for the left and right sides of this projector equation.

In turn, these should match the probabilities for both sides predicted by quantum theory, that is, expectations taken with respect to the system density operator. It seems, though, that the quantum theory prediction is being tested by the Peres experiment, and not any consequence of the deterministic HV rules. We are also not aware if such an experiment has been performed.

More generally, one could imagine experimentally challenging the sum rule, at the level of the valuation assignment itself, rather than at an expectation level as suggested by Peres. This was considered earlier by Glymour [32], in an extensive discussion of the sum rule in the context of general HV models. Glymour also proposed testing the assumption by reevaluating certain Compton scattering experimental outcomes, to test the rule, but we do not repeat his argument here. We are not aware if the suggested experiment has been done.

In any event, the Young double-slit scheme is a classic experiment. The reevaluation we described in Sec. V offers a clear, empirical no-go demonstration for the deterministic HV models. The simple spin-1 experiments described above have also been repeatedly performed: those outcomes only need to be reinterpreted in terms of quantum conditional probabilities to derive empirical no-go tests.

[1] S. Kochen and E. Specker, J. Math. Mech. **17**, 59 (1967).
[2] J. Bell, Rev. Mod. Phys. **38**, 447 (1966).
[3] A. Peres, J. Phys. A **24**, L175 (1991).
[4] A. Peres, Phys. Lett. A **163**, 243 (1992).
[5] N. D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990).
[6] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[7] D. Greenberger, M. Horne, A. Shimony, and A. Zeilinger, Am. J. Phys. **58**, 1131 (1989).
[8] A. Fine and P. Teller, Found. Phys. **8**, 629 (1978).
[9] A. Fine, Phys. Rev. Lett. **48**, 291 (1982).
[10] A. Fine, J. Math. Phys. **23**, 1306 (1982).
[11] A. Fine, *The Shaky Game: Einstein Realism and the Quantum Theory*, 2nd ed. (University of Chicago Press, Chicago, 1996).
[12] E. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1981).
[13] B. Arnold and S. Press, J. Am. Stat. Assoc. **84**, 152 (1989).
[14] B. van Fraassen, *Quantum Mechanics: An Empiricist Viewpoint* (Wiley, New York, 1991).
[15] J. Bub, *Interpreting the Quantum World* (Cambridge University Press, Cambridge, UK, 1997).
[16] S. Gudder, *Quantum Probability* (Academic, San Diego, 1988).
[17] M. Scully, H. Walther, and W. Schleich, Phys. Rev. A **49**, 1562 (1994).
[18] A. Bohm, *Quantum Mechanics. Foundations and Applications*, 3rd ed. (Springer-Verlag, New York, 1993).
[19] C. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
[20] J. Bub, Found. Phys. **3**, 29 (1973).
[21] J. Bub, Found. Phys. **6**, 511 (1976).
[22] S. Freedman and E. Wigner, Found. Phys. **3**, 457 (1973).
[23] J. Clauser, Am. J. Phys. **39**, 1095 (1971).
[24] J. Clauser, Am. J. Phys. **39**, 1098 (1971).
[25] E. Wigner, Am. J. Phys. **39**, 1097 (1971).
[26] S. Gudder, *Stochastic Methods in Quantum Mechanics* (North-Holland, New York, 1979).
[27] A. Agresti, *Introduction to Categorical Data Analysis* (Wiley, New York, 1996).
[28] Y. Bishop, S. Fienberg, and P. Holland, *Discrete Multivariate Analysis: Theory and Practice* (MIT Press, Cambridge, MA, 1975).
[29] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
[30] J. Devore, *Probability and Statistics for Engineering and the Sciences* (Brooks/Cole, Monterey, CA, 1982).
[31] L. Acardi, Phys. Rep. **77**, 169 (1981).
[32] C. Glymour, Philos. Sci. **44**, 86 (1977).